The generic differentiability of convex-concave functions: Characterization

Abbas Moameni *
Department of Mathematics and Statistics
Queen's University
Kingston, ON, Canada K7L3N6
momeni@mast.queensu.ca

January 12, 2011

Abstract

As established by R T. Rockafellar, real valued convex-concave functions are generically differentiable. It this paper we shall show that for a convex-concave function defined on an open convex set $C \times D$, there exist dense subsets $\mathcal{N}$ of C and $\mathcal{M}$ of D such that the partial derivative with respect to the first variable (resp. second variable) exists on $\mathcal{N} \times D$ (resp. $C \times \mathcal{M}$) and therefore the function is differentiable on $\mathcal{N} \times \mathcal{M}$. This is an interesting property of convex-concave functions and it does not hold for convex-convex functions. As an immediate application we recover the generic single-valuedness of monotone operators.

1 Introduction

Many results about generic differentiability of real valued convex functions are already known. However, to our knowledge the only results about directional derivatives and generic differentiability of real-valued convex-concave function was established by R T. Rockafellar in [8]. Since then there has been some contributions and extensions to the continuity and differentiability of convex-concave and biconvex operators taking vales in appropriate partially ordered vector spaces (see [4, 5] and references therein).

The regularity properties of convex-concave functions follows indeed as extensions of similar results for convex functions. That is why one may not expect to obtain a better result when it comes to convex-concave functions. Here is our main theorem that reveals an interesting property of convex-concave functions.

Theorem 1.1 *Let H be a convex-concave function on $\mathbb{R}^n \times \mathbb{R}^m$. Let $C \times D$ be an open convex set on which H is finite. The following statements hold,*
(1) There exists a dense subset $\mathcal{N}$ of C, such that $\mathcal{L}^n(C \setminus \mathcal{N}) = 0$ and for each $x \in \mathcal{N}$, the partial derivative $\nabla_1 H(x, y)$ exists for all $y \in D$.
(2) There exists a dense subset $\mathcal{M}$ of D such that $\mathcal{L}^m(D \setminus \mathcal{M}) = 0$ and for each $y \in \mathcal{M}$ the partial derivative $\nabla_2 H(x, y)$ exists for all $x \in C$.
(3) The complement of $\mathcal{N} \times \mathcal{M}$ in $C \times D$ has Lebesgue measure zero in $\mathbb{R}^n \times \mathbb{R}^m$ and H is differentiable on the dense subset $\mathcal{N} \times \mathcal{M}$ of $C \times D$.

The above conclusion no longer holds for convex-convex functions, for instance the function $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $H(x, y) = |x - y|$ fails to have this property. In fact, for each dense subset $\mathcal{N}$ of $\mathbb{R}$, the partial derivative with respect to the first variable does not exist on the whole set $\mathcal{N} \times \mathbb{R}$.

Here is an immediate corollary of this Theorem to skew-symmetric functions, i.e. functions define on $\mathbb{R}^n \times \mathbb{R}^n$ with $H(x, y) = -H(y, x)$ for all $(x, y) \in Dom(H)$.

*Research supported by a Coleman fellowship at Queen's University.

Corollary 1.2 *Let H be a convex-concave skew-symmetric function on $\mathbb{R}^n \times \mathbb{R}^n$. Let $C \times C$ be an open convex set on which H is finite. There exists a dense subset $\mathcal{N}$ of C, such that $\mathcal{L}^n(C \setminus \mathcal{N}) = 0$ and for each $x \in \mathcal{N}$, the function H is differentiable at (x, x).*

An important application of this Corollary is the generic single-valuedness of monotone operators. Indeed, as shown by E. Krauss [6] one can associate to each maximal monotone operator $T : \mathbb{R}^n \to \mathbb{R}^n$, a skew-symmetric closed convex-concave function H_T such that

$$Tx = y \qquad \text{if and only if } (y, -y) \in \partial H_T(x, x),$$

where ∂H_T stands for sub-differential of convex-concave function introduced by T. R. Rockafellar. By the above Corollary H_T is almost every where differentiable on the diagonal of $Dom(T) \times Dom(T)$ (assuming $int(Dom(T)) \neq \emptyset$) and therefore $\partial H_T(x, x) = \nabla H_T(x, x)$ for a dense subset of $Dom(T)$ for which we have the operator T is almost everywhere single-valued. For a detailed proof in more general spaces, the interested reader is referred to [7] where a new and shorter proof of the Krauss result together with the extension of Theorem 1.1 to mappings on Asplund topological spaces are provided.

The proof of Theorem 1.1 consists of permanence properties of convex-concave functions established by T. R. Rockafellar [8] together with a fundamental but less-known result of Arzela [1, 2] in (1883/1884) providing a necessary and sufficient condition for the point wise limit of a sequence of real valued continuous functions on compact sets to be continuous (see Theorem 3.4 in the present paper for the statement).

In the next section we recall some preliminary definitions and results which will be of use in section 3 where Theorem 1.1 is proved.

2 Preliminaries

In this section we start by introducing the notations used throughout the paper and then recall some of the standard results for both convex and convex-concave functions.

As to notation: If $x, y \in \mathbb{R}^n$ then the inner product is denoted by $\langle x, y\rangle_{\mathbb{R}^n}$. Explicitly if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$

$$\langle x, y\rangle_{\mathbb{R}^n} = x_1 y_1 + .. + x_n y_n.$$

The norm of x is $\|x\| = \sqrt{\langle x, x\rangle}$. Lebesgue measure in R^n will be denoted by $\mathcal{L}^n$ and integrals with respect to this measure will be written as $\int f(x)\, dx$. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be any function. Let x be a point where f is finite. The directional derivative of f at x in the direction u is denoted by $Df(x)u$ and is defined to be

$$Df(x)u = \lim_{\lambda \to 0^+} \frac{f(x + \lambda u) - f(x)}{\lambda}$$

if it exists. If f is differentiable at x, the directional derivatives $Df(x)u$ are all finite and

$$Df(x)u = \langle \nabla f(x), u\rangle_{\mathbb{R}^n}, \qquad \forall u \in \mathbb{R}^n,$$

where $\nabla f(x)$ is the gradient of f at x.
Let us list some of the properties of directional derivatives of convex functions.

Theorem 2.1 *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper convex function. Let x be a point where f is finite. The following statements hold,*
(1) For each $u \in \mathbb{R}^n$, the difference quotient in the definition of $Df(x)u$ is a non-decreasing function of $\lambda > 0$, so that $Df(x)u$ exists and

$$Df(x)u = \inf_{\lambda > 0} \frac{f(x + \lambda u) - f(x)}{\lambda}. \tag{1}$$

(2) $Df(x)u$ is a positively homogeneous convex function of u, with

$$Df(x)u + Df(x)(-u) \geq 0 \qquad \forall u \in \mathbb{R}^n.$$

Definition 2.2 *Say that a mapping* $H : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, \infty]$ *is a saddle function provided it is convex in the first variable and concave in the second; more precisely, we require that the function* $x \to H(x, y)$ *(resp.* $y \to -H(x, y)$*) be a convex function for each fixed* $y \in \mathbb{R}^m$ *(respectively each* $x \in \mathbb{R}^m$*).*

Similarly, one can define the directional derivatives for a saddle function $H : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, \infty]$ as follows Let $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ be a point where H is finite. The one sided directional derivative of H at (x, y) in the direction $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ is defined to be the limit

$$DH(x, y)(u, v) = \lim_{\lambda \to 0^+} \frac{H(x + \lambda u, y + \lambda v) - H(x, y)}{\lambda}$$

if the limit exists. By Theorem 2.1, the directional derivatives

$$D_1 H(x, y)(u) = \lim_{\lambda \to 0^+} \frac{H(x + \lambda u, y) - H(x, y)}{\lambda}$$

and

$$D_2 H(x, y)(v) = \lim_{\lambda \to 0^+} \frac{H(x, y + \lambda v) - H(x, y)}{\lambda}$$

exist and if H is differentiable with respect to the first variable (resp. second variable) at (x, y) then

$$D_1 H(x, y)(u) = \langle \nabla_1 H(x, y), u \rangle_{\mathbb{R}^n} \qquad \big(resp. \quad D_2 H(x, y)(v) = \langle \nabla_2 H(x, y), v \rangle_{\mathbb{R}^m} \big)$$

for all $u \in \mathbb{R}^n$ (resp. $v \in \mathbb{R}^m$) where $\nabla_1 H(x, y)$ (resp. $\nabla_2 H(x, y)$) is the gradient of H with respect to the first variable (resp. second variable) at (x, y). This result is due to T. Rockafellar [8].

Theorem 2.3 *Let* H *be a convex-concave function on* $\mathbb{R}^n \times \mathbb{R}^m$*. Let* $C \times D$ *be an open convex set on which* H *is finite. Then for each* $(x, y) \in C \times D$*,* $DH(x, y)(u, v)$ *exists and is a finite positively homogeneous convex-concave function of* (u, v) *on* $\mathbb{R}^n \times \mathbb{R}^m$*. In fact,*

$$DH(x, y)(u, v) = D_1 H(x, y)(u) + D_2 H(x, y)(v).$$

3 Proof of Theorem 1.1

Fix $(x, y) \in C \times D$. For each $\lambda > 0$, define the following functions on $\mathbb{R}^n \times \mathbb{R}^m$.

$$H_\lambda(u, v) = \frac{H(x + \lambda u, y + \lambda v) - H(x, y)}{\lambda},$$

$$\tilde{H}_\lambda(u, v) = \frac{H(x + \lambda u, y + \lambda v) - H(x, y + \lambda v)}{\lambda},$$

$H^1_\lambda(u) = H_\lambda(u, 0)$ and $H^2_\lambda(v) = H_\lambda(0, v)$. We have the following result regarding the convergence of $H_\lambda, \tilde{H}_\lambda, H^1_\lambda$ and H^2_λ.

Proposition 3.1 *The following statements hold.*
(1)$H^1_\lambda(u)$ *converge uniformly to* $D_1 H(x, y)(u)$ *on compact subsets of* $\mathbb{R}^n$*.*
(2)$H^2_\lambda(v)$ *converge uniformly to* $D_2 H(x, y)(v)$ *on compact subsets of* $\mathbb{R}^m$*.*
(3) $H_\lambda(u, v)$ *converges uniformly to* $D_1 H(x, y)(u) + D_2 H(x, y)(v)$ *on compact subsets* $A \times B$ *of* $\mathbb{R}^n \times \mathbb{R}^m$*.*
(4) Let $u \in \mathbb{R}^N$*. If* $\nabla_1 H(x, y)$ *exists then for each* $v \in \mathbb{R}^m$*,*

$$\lim_{\lambda \to 0^+} D_1 H(x, y + \lambda v)(u) = D_1 H(x, y)(u),$$

and this convergence is uniform on compact subsets of $\mathbb{R}^m$*.*
(5) $\tilde{H}_\lambda(u, v)$ *converges uniformly to* $D_1 H(x, y)(u)$ *on compact subsets* $A \times B$ *of* $\mathbb{R}^n \times \mathbb{R}^m$*.*

We shall need the following result known as Dini's theorem for the proof of this Proposition.

Theorem 3.1 *Let K be a compact subset in a metric space, and*
(1) $\{f_k\}$ is a sequence of continuous functions on K,
(2) $\{f_k\}$ converges pointwise to a continuous function f on K,
(3) $f_k(x) \geq f_{k+1}(x)$ for all $x \in K, k = 1, 2, 3, ...$
Then $f_k \to f$ uniformly on K and therefore f is continuous on K.

Proof of Proposition 3.1. For parts (1) and (2), note that H^1_λ is a non-decreasing function and H^2_λ is a non-increasing function of $\lambda > 0$ and therefore the result follows from Dini's Theorem.
Proof of part (3): We first show that for each $\epsilon > 0$ there exists $\lambda_0 > 0$ such that for all $0 < \lambda < \lambda_0$, we have

$$H_\lambda(u, v) < D_1H(x, y)(u) + D_2H(x, y)(v) + \epsilon, \quad \text{for all } (u, v) \in A \times B.$$

Then by a dual argument we have

$$H_\lambda(u, v) > D_1H(x, y)(u) + D_2H(x, y)(v) - \epsilon, \quad \text{for all } (u, v) \in A \times B,$$

from which we obtain the desired result in part (3). The difference quotient in the function H_λ can be expressed as

$$\frac{H(x, y + \lambda v) - H(x, y)}{\lambda} + \frac{H(x + \lambda u, y + \lambda v) - H(x, y + \lambda v)}{\lambda},$$

where the first quotient converges uniformly to $D_2H(x, y)(v)$ on B. It follows from Theorem 35.1 in [8] that H is Lipschitz on every closed bounded set of $C \times D$. Suppose $\Lambda > 0$ is the Lipschitz constant on the set

$$K := \{(x + \delta_1 A) \times (y + \delta_2 B); 0 \leq \delta_1, \delta_2 \leq \delta_0\}$$

where $\delta_0 > 0$ is small enough so that $K \subset C \times D$. Since $H^1_\lambda(u)$ converges uniformly to $D_1H(x, y)(u)$ on A, there exists $0 < \alpha < \delta_0$ such that

$$\frac{H(x + \alpha u, y) - H(x, y)}{\alpha} < D_1H(x, y)(u) + \frac{\epsilon}{2}.$$

Thus, for every $v \in B$ and $0 < \lambda < \delta_0$, it follows from the above inequality that

$$\frac{H(x + \alpha u, y + \lambda v) - H(x, y + \lambda v)}{\alpha} < D_1H(x, y)(u) + \frac{\epsilon}{2} + \frac{2\lambda\Lambda\|v\|}{\alpha}.$$

Let λ_0 be small enough such that $\frac{2\lambda_0\Lambda\|v\|}{\alpha} < \epsilon/2$ for all $v \in B$. For each $0 < \lambda < \min\{\lambda_0, \alpha\}$ we have

$$\begin{aligned} D_1H(x, y)(u) + \epsilon &> \frac{H(x + \alpha u, y + \lambda v) - H(x, y + \lambda v)}{\alpha} \\ &\geq \frac{H(x + \lambda u, y + \lambda v) - H(x, y + \lambda v)}{\lambda}, \end{aligned} \tag{2}$$

from which part (3) follows.
We know prove part (4). Note first that by (1) in Theorem 2.1 we have $\frac{H(x+\lambda u,y+\lambda v)-H(x,y+\lambda v)}{\lambda} \geq D_1H(x, y+\lambda v)(u)$, from which together with (2) we have

$$D_1H(x, y)(u) + \epsilon > D_1H(x, y + \lambda v)(u), \tag{3}$$

for every $0 < \lambda < \min\{\lambda_0, \alpha\}$ and $v \in B$. By a similar argument we have

$$-D_1H(x, y)(-u) - \epsilon < -D_1H(x, y + \lambda v)(-u), \tag{4}$$

It follows from (3), (4) and part (2) of Theorem 2.1 that

$$-D_1H(x, y)(-u) - \epsilon < -D_1H(x, y + \lambda v)(-u) \leq D_1H(x, y + \lambda v)(u) < D_1H(x, y)(u) + \epsilon,$$

and the result follows due to assumption that $\nabla_1H(x, y)$ exists and $D_1H(x, y)(u) = -D_1H(x, y)(-u) = \langle\nabla_1H(x, y), u\rangle_{\mathbb{R}^n}$.
Part (5) follows from the fact that $\tilde{H}_\lambda(u, v) = H_\lambda(u, v) - H^2_\lambda(v)$. □

The following result holds for both convex-concave and convex-convex functions.

Proposition 3.2 *Suppose F is a finite convex-concave (resp. convex-convex) function on convex subset $C \times D$ of $\mathbb{R}^n \times \mathbb{R}^m$. The following assertions hold:*
(1) There exists a dense subset $\mathcal{N}$ of C, such that $\mathcal{L}^n(C \setminus \mathcal{N}) = 0$ and for each $x \in \mathcal{N}$, the partial derivative $\nabla_1 H(x,y)$ exists for almost all $y \in D$.
(2) There exists a dense subset $\mathcal{M}$ of D such that $\mathcal{L}^m(D \setminus \mathcal{M}) = 0$ and for each $y \in \mathcal{M}$ the partial derivative $\nabla_2 H(x,y)$ exists for almost all $x \in C$.

Proof. We just prove part (1). Part (2) follows from a similar argument. For each $x \in C$, set

$$N_x = \{y \in D; \nabla_1 H(x,y) \quad \text{does not exist } \}.$$

Thus, the set $N = \cup_{x\in C}\{x\} \times N_x$ consists of all $(x,y) \in C \times D$ such that $\nabla_1 H(x,y)$ does not exist. It follows that for convex-concave (resp. convex-convex) functions this set is of measure zero in $\mathbb{R}^n \times \mathbb{R}^m$. Therefore, it follows from Fubini's theorem that

$$0 = \iint\limits_N dy\, dx = \int_C \int_D 1_{\{x\}\times N_x}\, dy\, dx = \int_C \mathcal{L}^m(N_x)\, dx.$$

Therefore, there exists a dense subset $\mathcal{N}$ of C, such that $\mathcal{L}^n(C \setminus \mathcal{N}) = 0$, and for each $x \in \mathcal{N}$ one has $\mathcal{L}^m(N_x) = 0$ from which one has $D \setminus N_x$ is dense in D. □

Lemma 3.2 *Let $\mathcal{N}$ and $\mathcal{M}$ be as in Proposition 3.2 for the convex-concave function H in Theorem 1.1. The following statements hold.*
(1) Let $x_0 \in \mathcal{N}$. For each $u \in \mathbb{R}^n$ the function $f_u : D \to \mathbb{R}$ defined by $f_u(y) = D_1H(x_0, y)(u)$ is continuous.
(2) Let $y_0 \in \mathcal{M}$. For each $v \in \mathbb{R}^m$ the function $g_v : C \to \mathbb{R}$ defined by $g_v(x) = D_2H(x, y_0)(v)$ is continuous.

Note that the functions f_u and g_v in the above Lemma are indeed the pointwise limit -not necessary uniform though- of the quotients in the definition of directional derivatives of the function H. To prove the continuity of these functions we first recall the notion of *quasi-uniformly* convergence and an immediate application introduced by Arzela [1, 2].

Definition 3.3 *A sequence $\{f_k\}$ of (scalar-valued) functions on an arbitrary set X is said to converge to f quasi-uniformly on X, if $\{f_k\}$ converges pointwise to f and if, for every $\epsilon > 0$ and $L \in \mathbb{N}$, there exists a finite number of indices $k_1, k_2, ..., k_l \geq L$ such that for each $x \in X$ at least one of the following inequalities holds:*

$$|f_{k_i}(x) - f(x)| < \epsilon, \qquad i = 1, 2, ..., l.$$

Theorem 3.4 *If a sequence of real-valued functions on a topological space X converges to a concinnous limit, then the convergence is quasi-uniform on every compact subset of X. Conversely, if the sequence converges quasi-uniformly on a subset of X, the limit is continuous on this subset.*

The interested reader is also referred to a more recent paper [3] for the proof.

Proof of Lemma 3.2. We just prove part (1). A similar argument yields part (2). Let B be a compact subset of D with a non-empty interior. Fix $u \in \mathbb{R}^n$. We shall show that $y \to f_u(y) = D_1H(x_0, y)(u)$ is continuous on B. In order to simplify the writing and since the direction u is fixed, we do not indicate the dependence of the function f_u to u and we just use f instead of f_u.
Note first that $f(y) = D_1H(x_0, y)(u) = \lim_{k\to\infty} f_k(y)$ where

$$f_k(y) = \frac{H(x_0 + \lambda_k u, y) - H(x_0, y)}{\lambda_k},$$

and $\lambda_k = 1/k$ for $k \in \mathbb{N}$. Note that for each k the function f_k is continuous. We shall show that f_k converges quasi-uniformly to f on B. Fix $\epsilon > 0$ and $L \in \mathbb{N}$. Since $x_0 \in \mathcal{N}$, it follows from Proposition 3.2 that there

exists a dense subset B_{x_0} of D such that $\nabla_1 H(x_0,y)$ exists for all $y\in B_{x_0}$. For each $y\in B_{x_0}$, it follows from parts (4) and (5) of Proposition 3.1 that there exists $k_y>L$ such that

$$\Big|\frac{H(x_0+\lambda_{k_y}u,y+\lambda_{k_y}v)-H(x_0,y+\lambda_{k_y}v)}{\lambda_{k_y}}-D_1H(x_0,y)(u)\Big|<\frac{\epsilon}{2},$$

and

$$\big|D_1H(x_0,y+\lambda_{k_y}v)(u)-D_1H(x_0,y)(u)\big|<\frac{\epsilon}{2},$$

for every $v\in B$. Note that $f_{k_y}(y+\lambda_{k_y}v)=[H(x_0+\lambda_{k_y}u,y+\lambda_{k_y}v)-H(x_0,y+\lambda_{k_y}v)]/\lambda_{k_y}$ and $f(y+\lambda_{k_y}v)=D_1H(x_0,y+\lambda_{k_y}v)(u)$. Thus, it follows from the above inequalities that

$$|f_{k_y}(y+\lambda_{k_y}v)-f(y+\lambda_{k_y}v)|<\epsilon, \tag{5}$$

for all $v\in B$. Define $U_y=\{y+\lambda_{k_y}v; v\in B\}$. Since B_{x_0} is dense in D we have

$$B\subset\cup_{y\in B_{x_0}}int\big(U_y\big).$$

B is compact and therefore there exist $y_1,y_2,...,y_l\in B_{x_0}$ such that $B\subset\cup_{i=1}^{l}int\big(U_{y_i}\big)$. This together with (5) implies that

$$|f_{k_{y_i}}(w)-f(w)|<\epsilon,\qquad \text{for all } w\in U_{y_i},$$

and therefore f_k converges to f quasi-uniformly on B from which we have f is continuous on B. $\square$

Proof of Theorem 1.1. Let $\mathcal{N}$ and $\mathcal{M}$ be as in Lemma 3.2. Fix $x_0\in\mathcal{N}$. We shall show that $\nabla_1H(x_0,y)$ exists for all $y\in D$. It follows from Proposition 3.2 there exists a dense subset B_{x_0} of D such that $\nabla_1H(x_0,y)$ exists for all $y\in B_{x_0}$. Fix $u\in\mathbb{R}^N$. We need to show that

$$D_1H(x_0,y)(u)+D_1H(x_0,y)(-u)=0, \tag{6}$$

for all $y\in D$. Note first that, equality (6) holds for all $y\in B_{x_0}$. It also follows from Lemma 3.2 that the function $y\to D_1H(x_0,y)(u)+D_1H(x_0,y)(-u)$ is continuous and therefore the result follows from the density of B_{x_0} in D.
By a similar argument we have that for every $y_0\in\mathcal{M}$, $\nabla_2H(x,y_0)$ exists for all $x\in C$. It finally follows that $\mathcal{N}\times\mathcal{M}$ is dense in $C\times D$ and for each $(x_0,y_0)\in\mathcal{N}\times\mathcal{M}$, the function H is differentiable at (x_0,y_0). $\square$

References

[1] C. Arzelà, *Intorno alla continuità della somma d'infinità di funzioni continue,* Rend. R. Accad. Sci. Istit. Bologna (1883/1884) 79-84.

[2] C. Arzelà, *Sulle serie di funzioni,* Mem. R. Accad. Sci. Ist. Bologna 5 (8) (1899/1900) 131-186.

[3] R.G. Bartle, *On compactness in functional analysis,* Trans. Amer. Math. Soc. 79 (1955) 3557.

[4] J. M. Borwein, *Partially monotone operators and the generic differentiability of convex-concave and biconvex mappings.* Israel J. Math. 54 (1986), no. 1, 42-50.

[5] M. Jouak, L. Thibault, *Directional derivatives and almost everywhere differentiability of biconvex and concave-convex operators.* Math. Scand. 57 (1985), no. 1, 215224.

[6] E. Krauss *A representation of arbitrary maximal monotone operators via subgradients of skew-symmetric saddle functions.* Nonlinear Anal. 9 (1985), no. 12, 1381-1399,

[7] A. Moameni *A charactrization of points of differentiability of convex-concave mappings,* In prepration.

[8] T. Rockafellar, *Convex Analysis,* 1970, Princeton University Press.